

Durham Research Online

Deposited in DRO:

22 November 2010

Version of attached file:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Inglis, M. and Simpson, A. (2008) 'Conditional inference and advanced mathematical study.', Educational studies in mathematics., 67 (3). pp. 187-204.

Further information on publisher's website:

<http://dx.doi.org/10.1007/s10649-007-9098-9>

Publisher's copyright statement:

NOTICE: this is the author's version of a work that was accepted for publication in Educational studies in mathematics.

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

Conditional inference and advanced mathematical study[†]

Matthew Inglis

Learning Sciences Research Institute, University of Nottingham.

Adrian Simpson

School of Education, University of Durham.

Abstract. Many mathematicians and curriculum bodies have argued in favour of the theory of formal discipline: that studying advanced mathematics develops one's ability to reason logically. In this paper we explore this view by directly comparing the inferences drawn from abstract conditional statements by advanced mathematics students and well-educated arts students. The mathematics students in the study were found to endorse fewer invalid conditional inferences than the arts students, but they did not endorse significantly more valid inferences. We establish that both groups tended to endorse more inferences which led to negated conclusions than inferences which led to affirmative conclusions (a phenomenon known as the negative conclusion effect). In contrast, however, we demonstrate that, unlike the arts students, the mathematics students did not exhibit the affirmative premise effect: the tendency to endorse more inferences with affirmative premises than with negated premises. We speculate that this latter result may be due to an increased ability for successful mathematics students to be able to 'see through' opaque representations. Overall, our data are consistent with a version of the formal discipline view. However, there are important caveats; in particular, we demonstrate that there is no simplistic relationship between the study of advanced mathematics and conditional inference behaviour.

Keywords: advanced mathematical thinking, conditional inference, logic, reasoning, representation systems, theory of formal discipline

One of the original rationales for placing mathematics at the heart of the school curriculum was the idea that studying mathematics develops one's abstract logical thinking skills (e.g. Davis, 1970). This is a variant of what Thorndike (1924) called the *Theory of Formal Discipline*, the idea that the study of certain disciplines benefits reasoning generally. Oakley (1946) typified this view, writing that "the study of mathematics cannot be replaced by any other activity that will train and develop man's purely logical faculties to the same level of rationality." (p.19). Similar views are still widespread today: when discussing the utility of studying undergraduate level mathematics, the QAA (2002) – the UK quality assurance agency for higher education – stated that "[mathematics] graduates are rightly seen as possessing considerable skill in abstract reasoning, logical deduction and problem solving". Comparable arguments have been made at the school level. In a report on post-14 mathematics in the UK, Smith (2004) argued that "mathematical training disciplines the mind, develops logical and critical reasoning, and develops analytical and problem-solving skills to a high degree" (p.11).

[†] This is a preprint of a paper which appeared in *Educational Studies in Mathematics*, 67(3), 187-204. The final version is available at: <http://dx.doi.org/10.1007/s10649-007-9098-9>

These widely held views appear to contradict the beliefs of early psychologists. Thorndike and Woodworth (1901), for example, had found that “improvement in any single mental function rarely brings about equal improvement in any other function, no matter how similar” (p. 250). Instead they suggested that studying abstract systems such as mathematics or Latin has no impact upon students’ logical reasoning skills (Thorndike, 1924). However, in a later study of science, humanities and psychology undergraduates, Lehman and Nisbett (1990) found a moderate correlation ($r = 0.31$) between the number of mathematics courses a student took during their studies and their improvement on a conditional inference test (taken once at the start of their degree course, and once at the end). Conditional inference (drawing conclusions on the basis of statements of the form ‘if p then q ’ and a premise) is clearly fundamental to formal logical reasoning, and Lehman and Nisbett suggested their findings provided some support for the theory of formal discipline.

Establishing the veracity of the formal discipline view is not only of theoretical interest, it has potentially important policy implications. In a survey of early twentieth century US mathematics curricula, Stanic (1986) noted that changes to curricula were substantially related to changing attitudes towards the theory of formal discipline. More recently, Smith (2004) cited the formal discipline view as evidence for the utility of mathematics in a report to the UK government. Amongst other policy recommendations which followed from this conclusion, Smith advocated implementing university tuition fee waivers for mathematics students, and targeted salary increases for school mathematics teachers. In short, firmly establishing the accuracy (or inaccuracy) of the formal discipline theory could have clear educational policy implications and it should be regarded as an important goal of mathematics education research.

Surprisingly, given the importance and long history of these ideas – which date back at least as far as Plato – there have apparently been no previous studies which have directly and systematically interrogated the theory of formal discipline by comparing the conditional inference behaviour of advanced mathematics students with those from other disciplines. The primary goal of this paper then, is to fill that gap by directly comparing how successful mathematics and arts students draw inferences from abstract conditional statements.

1. Conditional inference and mathematics.

Given a statement ‘if p then q ’ there are four inferences which are commonly drawn, only two of which are logically valid; the different inferences are summarised in Table I. In standard logic, the *modus ponens* (MP) and *modus tollens* (MT) inferences are valid, whereas the *denial of the antecedent*

Table I. The four inferences.

	MP	DA	AC	MT
Conditional	if p then q	if p then q	if p then q	if p then q
Premise	p	not- p	q	not- q
Conclusion	q	not- q	p	not- p
Type	affirmative	denial	affirmative	denial
Validity	valid	invalid	invalid	valid

(DA) and *affirmation of the consequent* (AC) inferences are invalid.¹ Those inferences which, from the rule ‘if p then q ’, lead to positive conclusions (i.e. conclusions without negated components) are known as ‘affirmative’ inferences, whereas those which lead to negative conclusions are referred to as ‘denial’ inferences. Despite their invalidity, DA and AC inferences are often made in day-to-day life. For example when a political party announces that “if our opponents win, interest rates will rise”, they are hoping that the electorate will make the DA inference: that voters will infer that if they win, interest rates will not rise.

1.1. CONDITIONAL INFERENCE IN MATHEMATICS EDUCATION.

While there are other skills involved in logical reasoning performance, conditional inference is clearly an important component. Given this, it is unsurprising that a number of educational researchers have focussed on how conditionals are dealt with by secondary school students (Hoyles and Küchemann, 2002; O’Brien et al., 1971; O’Brien, 1972), undergraduates (Durand-Guerrier, 2003; O’Brien, 1973; Stylianides et al., 2004), and teachers (Damarin, 1977; Eisenberg and McGinty, 1974). In this section we briefly review the main findings of those studies which have looked at university students’ conditional inference behaviour.

O’Brien (1973) studied the conditional inference behaviour of undergraduates taking an introduction to mathematics course. He gave his participants a series of conditionals set in different contexts (e.g. “if Sue screams, Jim will jump”) together with a minor premise (“Jim did not jump”) and asked them whether inferences could be made (“did Sue scream?”). O’Brien found that only 3 participants (from a sample of 60) reasoned normatively across all the tasks. One particular source of difficulty was the MT inference: across

¹ DA and AC are not the only invalid inferences participants can draw. For example, ‘if p then q , p , therefore not- q ’ is a further possible invalid inference. No previous study has shown evidence that such inferences are widely drawn and we follow the conventions set in the literature by not considering them.

the study only 57% of these inferences were endorsed, compared to nearly 100% for MP. Importantly, O'Brien found that the frequency that the MT inference was drawn by depended upon the context in which the rule was situated. Instead of behaving normatively, approximately half of O'Brien's (1973) participants used what he termed 'child logic': they consistently drew all the inferences: MP, DA, AC and MT. O'Brien accounted for this by suggesting that many college students interpret an 'if p then q ' conditional as a ' p if and only if q ' biconditional. However Durand-Guerrier (2003), using conditionals set in various different mathematical contexts, found that few students consistently interpreted the conditional in this fashion. Instead, she argued that whether this interpretation was made related to the participant's background knowledge of the context in which the particular conditional was set.

In a later study, Stylianides et al. (2004) investigated the differences between mathematics undergraduates' and education undergraduates' behaviour when drawing DA and MT inferences from contextualised conditionals (in what they called a 'verbal context'). They also presented participants with a proof validation task to assess MT inferences in a 'symbolic context' (in our terms, both these versions were contextual, as responses could have been influenced by prior knowledge). They concluded that, when drawing conditional inferences, participants' mathematical backgrounds interact with the context in which the conditional is set.

There are, however, several difficulties with Stylianides et al.'s (2004) study which render it hard to interpret their results. First, the control group of education students who participated in the study had apparently taken several core proof-based undergraduate mathematics modules which had emphasised logical thinking. Consequently, in the absence of a genuine control group, it would be hard to draw conclusions from the study about the relationship between logical thinking and the study of advanced mathematics. Second, the two contexts used in the study – verbal and symbolic – were studied using non-isomorphic tasks: one used a conditional inference task² and one used a proof validation task. It would be hard, therefore, to conclude that any between-contexts differences were due to factors related to context rather than to the task structure. Third, the dichotomous data reported in the paper appear to have been analysed using parametric statistical methods. Nevertheless, Stylianides et al.'s observation that the context in which a conditional statement is situated affects how participants draw inferences from it is important and certainly correct (e.g. Newstead et al., 1997). Given this, it should

² To further complicate interpretation, two different rule types were used in the verbal context task. For the MT inference a rule of the form 'if p then q ' was used, whereas the DA inference used an 'if not- p then q ' rule. As the current study demonstrates, the presence or absence of negated components in conditional statements has a large effect upon whether inferences are made.

be noted that neither O'Brien (1973), Durand-Guerrier (2003) or Stylianides et al. used genuinely abstract materials (in the sense that participants' responses could not have been influenced by their prior knowledge of the context).

1.2. CONDITIONAL INFERENCE IN PSYCHOLOGY

Given the long history of the formal discipline view, and given the finding that the context in which a task is set interferes with the analysis of participants' underlying logical behaviour, it is perhaps surprising that no previous study has directly compared the conditional inferences that are drawn by advanced mathematics students and students of other disciplines, from purely abstract materials (i.e. with materials far removed from a specific context).

There have, however, been numerous psychological studies on conditional inference using such materials. These studies have tended to involve social science undergraduates, and much theory has been built up to account for the data resulting from them (e.g. Evans, 2007; Johnson-Laird, 2006; Oaksford and Chater, 2007). Given the proliferation of theories in this domain, it is important to emphasise that our goal in this paper is not to intervene in debates between the various reasoning theorists, instead we seek to study the relationship between conditional inference behaviour and the advanced study of mathematics.

Typically an abstract conditional inference task in these studies involves a series of imagined letter-number pairs. Participants are given a rule such as "if the letter is D then the number is 4", together with a premise of the form "the number is not 4", and are asked to decide whether or not a conclusion (e.g. "the letter is not D") necessarily follows. An example of a trial from a conditional inference task is shown in Figure 1. In the so-called negations paradigm the task is varied by rotating the position of negatives in the rule giving a total of 16 possible inferences. These possibilities are summarised in Table II. Two main effects which hinder logical behaviour have been identified in existing studies which use this experimental paradigm: the *negative conclusion effect* and the *affirmative premise effect*.³

The negative conclusion effect refers to the finding that participants tend to draw more inferences with negative conclusions than they do inferences with affirmative conclusions. That is to say that the inference 'if A then 3; not-3; therefore not-A' is made more often than the inference 'if not-A then 3; not-3; therefore A' despite, in this case, both being valid MT inferences. This

³ Due to differing theoretical interpretations of these effects some researchers have used different terminology. The negative conclusion effect is also sometimes referred to as negative conclusion bias, or as the double negation effect. Similarly, the affirmative premise effect has been called affirmative premise bias, or the implicit negation effect (e.g. Evans and Handley, 1999; Oaksford et al., 2000). We have attempted to adopt theoretically neutral terminology.

This problem concerns an imaginary letter-number pair. Your task is to decide whether or not the conclusion *necessarily* follows from the rule and the premise.

Rule: If the letter is not H then the number is 1.

Premise: The number is not 1.

Conclusion: The letter is H.

☐ YES (it follows) ☐ NO (no, it does not follow)

Figure 1. A typical conditional inference task (for the rule ‘if not- p then q ’, the inference MT, and with an explicitly negated premise ‘not- q ’).

Table II. The four inferences with and without negated premises (Pr) and conclusions (Con).

Conditional	MP		DA		AC		MT	
	Pr	Con	Pr	Con	Pr	Con	Pr	Con
if p then q	p	q	not- p	not- q	q	p	not- q	not- p
if p then not- q	p	not- q	not- p	q	not- q	p	q	not- p
if not- p then q	not- p	q	p	not- q	q	not- p	not- q	p
if not- p then not- q	not- p	not- q	p	q	not- q	not- p	q	p

effect has been robustly found on both denial inferences (DA and MT), but is only weakly observed (if at all) on AC, and never on MP (Schroyens et al., 2001). One account of this surprising finding maintains that it is caused by the difficulty participants have with the logical equivalence between not-(not- p) and p (Evans et al., 1995; Evans and Handley, 1999; Schroyens et al., 2000). This is because on both denial inferences, affirmative conclusions require participants to take one extra logical step. To clarify, consider making an MT inference from the two conditionals above: (i) ‘if A then 3’ and (ii) ‘if not-A then 3’. Given the premise not-3, to draw a valid inference with statement (i) requires drawing one logical inference (MT). But given the same premise, reaching a valid inference with statement (ii) requires an extra step: MT to deduce not-(not-A) and then an application of the equivalence of not-(not- p) and p .⁴

The affirmative premise effect refers to the finding that participants tend to endorse more inferences from affirmative premises than from negative premises. It tends to primarily be observed when those negative premises are represented implicitly. That is to say that the inference ‘if not-A then

⁴ An alternative account for the negative conclusion effect states that it is a consequence of a preconscious bias in favour of negative conclusions (Pollard and Evans, 1980). Oaksford et al. (2000) pointed out that, for any p , the probability of not- p tends to be greater than that of p . Consequently, they suggested that there could be an inbuilt cognitive reluctance to draw positive inferences, as they are more likely to be wrong (although see Schroyens et al., 2000).

3; A; therefore not-3' is made more often than the inference 'if A then 3; R; therefore not-3', even though they are both instances of drawing, invalidly, the DA inference. Evans and Handley (1999) suggested that this was caused by the difficulty of recognising that the premise was relevant to the conditional when it is implicitly negated. So the premise 'R' is less obviously 'about' the conditional 'if A then 3' than the premise 'A' is 'about' the conditional 'if not-A then 3'. This account is discussed in greater depth later in the paper.

Evans and Handley (1999) brought the negative conclusion and affirmative premise effects together in what they termed the 'two hurdle' account. They noted that if a conditional inference is to be made two hurdles need to be jumped. Firstly, the reasoner must see that the given premise is relevant to the conditional statement. If this hurdle is passed, and the reasoner believes that a legitimate inference can be made, on some occasions there is a further hurdle: the reasoner must be able to convert statements of the form not-(not- p) into the statement p . It is only if both hurdles are successfully traversed that an inference can be made.

In sum, several mathematics education studies have investigated mathematics students' conditional inferential behaviour on contextual materials, and several psychological studies have looked at the conditional inference behaviour of well-educated but non-mathematical participants on abstract materials. No study, however, has considered the conditional inference behaviour of mathematics students using abstract materials and, in particular, whether this might differ from non-mathematical participants. This is a curious omission, for two reasons. First, the studies which have looked at the contextual conditional reasoning of mathematics students have found that the context is a confounding factor; so if one wishes to study the underlying logical behaviour of participants it would be prudent to strip away as many extraneous contextual factors as possible. Second, many commentators and curriculum bodies have argued for a version of the formal discipline view, by suggesting that studying mathematics at an advanced level develops students' abilities for abstract conditional reasoning. Despite this claim being widely made, there have been few empirical studies which have directly investigated it, and none have directly compared mathematics and non-mathematics students' abstract conditional inference behaviour.

Specifically, then, the goal of this study was to examine the formal discipline theory via two main research questions: (i) Do advanced mathematics students draw more normative inferences from abstract conditional statements than students who have not studied advanced mathematics? (ii) Do successful mathematics students exhibit either (a) the negative conclusion effect or (b) the affirmative premise effect? If Thorndike and Woodworth (1901) and the early psychologists were right, and a training in mathematics has no effect upon abstract logical reasoning skills, one would expect there to be few differences between the extent to which mathematics and non-

mathematics students are affected by the negative conclusion and affirmative premise effects. If, however, the theory of formal discipline (and Lehman and Nisbett's (1990) findings in support of it) is correct, and if studying mathematics does develop abstract logical reasoning skills, one might expect mathematics students to exhibit reduced negative conclusion and affirmative premise effects.

2. Method.

When designing a study to answer the research questions set out above, several factors are relevant. Of special concern is that, by most accounts, a large component of conditional reasoning behaviour is determined by pre-conscious cognitive processes (e.g. Evans, 2007). Indeed, by some accounts, *all* reasoning behaviour is determined in this way (e.g. Oaksford and Chater, 2007). Preconscious cognitive processes are, by definition, unavailable for participants to report upon concurrently, and so verbal protocol analyses may be at best unhelpful, and at worst misleading (Nisbett and Wilson, 1977; Wason and Evans, 1975). Further, there is some evidence that asking participants to 'think aloud' when tackling tasks which have a large non-conscious component may detrimentally interfere with their success rates (Schooler et al., 1993). Given these factors, we adopted a purely quantitative task-based approach, and adapted the materials and procedure used by Evans et al. (1995).

Each participant was presented with 32 reasoning problems of the form shown in Figure 1. Half of the questions pertained to the inferences shown in Table II, whereas the remaining half were identical in form, except where explicitly negated premises had been replaced with implicitly negated premises (i.e. premises such as 'not-3' were replaced with, for example, '7'). The lexical content of the rules were generated randomly, and the order of the problems was randomised for each participant. Participants were tested in groups without time restriction, and were presented with a problem book with six pages of problems, and a page of instructions – shown in Figure 2 – adapted from those used by Evans et al. (1995).

Participants were 65 paid volunteers, consisting of 34 arts and social science (henceforth arts) undergraduates, and 31 mathematics undergraduates (on either single honours programmes, or joint honours programmes which contained a substantial proof-based mathematics component). None of the arts students were on programmes of study which involved taking proof-based mathematics modules. The mathematics students, in contrast, had all taken a compulsory module on the Foundations of Mathematics (based on Johnson, 1998). Along with sections on set theory, functions and cardinality, the course contained several hours of formal teaching about mathematical

This experiment is concerned with people's ability to reason logically with sentences in various forms. You will be presented with a total of 32 problems on the attached pages. In each case you are given two statements together with a conclusion which may or may not follow from these statements. Your task in each case is to decide whether or not the conclusion *necessarily* follows from the statements. A conclusion is necessary if it must be true, given that the statements are true. Each problem concerns an imaginary letter-number pair and contains an initial statement or rule which determines which letters may be paired with which numbers. An example of a rule of similar form to those used would be:

If the letter is B then the number is not 7.

In each case you must assume that the rule holds and then combine it with the information given in the second statement. This will concern either the letter or the number of an imaginary pair, for example:

The letter is Y.

The number is not 4.

If the information concerns the letter the conclusion will concern the number and vice-versa.

A full problem looks something like this:

If the letter is X then the number is 1.

The letter is X.

Conclusion: The number is 1. ☐ YES ☐ NO

If you think the conclusion necessarily follows please tick the YES box, otherwise tick the NO box. The experiment consists of this instructions page, and six pages of problems. Please work through the problems in order and make sure you do not miss any. Do not return to a problem once you have finished and moved on to another.

Figure 2. The instruction page of the booklet given to participants.

logic, including universal and existential quantification, contradiction proofs, contrapositive arguments and so on. (Note that there is no reason to believe that this course, in itself, would make students more likely to reason normatively: Cheng et al. (1986) found that attending a complete term of lectures on logic had no effect on (non-mathematics) students' performance on logical reasoning tasks).

3. Results.

We present the results of this study in three sections. First we discuss the overall inferential behaviour of the participants; second we look at the negative conclusion effect; and finally we consider the affirmative premise effect.

3.1. OVERALL RESPONSES.

The number of normatively correct responses was calculated for each participant: these ranged from 12 (41%) to 32 (100%). As shown in Table III, the mathematical group had a higher mean number of normative answers (Mean = 26.6, Std. Dev. = 4.1) than did the arts group (Mean =

Table III. The mean number of each inference made by each group, with the statistical significance of the differences (t -tests). * difference significant at the 0.001 level.

	MP	DA	AC	MT	total correct
Arts	7.56	3.88	5.68	4.91	18.91
Maths	7.84	1.03	1.77	5.52	26.55
t	1.93	5.29*	7.12*	1.21	7.23*

18.9, Std. Dev. = 4.4), a difference significant at the 0.001 level. Recalling that each type of inference occurred 8 times in the test, considering each type separately indicates where this difference originates. There was no significant difference between the mean number of MP or MT inferences made by each group: MP inferences were almost always drawn by both groups, whereas MT inferences were drawn around two-thirds of the time for each group. In contrast, the mathematics students made significantly fewer invalid (DA and AC) inferences than the arts students, both $ps < 0.001$. In line with Durand-Guerrier's (2003) observation that few students consistently behave in accordance with what O'Brien (1973) termed 'child logic', we found that only one participant (an arts student) endorsed every inference.

3.2. THE NEGATIVE CONCLUSION EFFECT.

The number of inferences made by participants with negative and affirmative conclusions is shown in Table IV. To conduct the analysis we calculated a *Negative Conclusion Index* (NCI), defined by the number of inferences endorsed on arguments with negative conclusions minus the number of inferences endorsed on arguments with affirmative conclusions. Thus, since the number of each type of both valid and invalid inferences with negative and affirmative conclusions was equal, a participant exhibiting no negative conclusion effect would be expected to attain an overall NCI of zero, and a participant who did exhibit the effect would be expected to have a positive NCI. For each group the significance of the negative conclusion effect was assessed by comparing the NCIs to zero using a one-sample Wilcoxon signed ranks test.⁵ Overall, both groups showed a strong effect (both $ps < 0.001$). A similar analysis was conducted for each type of inference in turn. The arts group showed a significant effect on the denial MT, $p < 0.001$, and DA, $p = 0.004$, inferences; but no significant effects on either MP or AC.

⁵ In view of the large number of comparisons (ten) being made in this set of Wilcoxon signed-rank analyses, we set a more stringent α -value of 0.01, rather than the normal 0.05.

Table IV. The percentage of inferences with negative and affirmative conclusions made by each group (Art – arts, Mat – mathematics), collapsed across premise-type (explicit/implicit negatives).

	MP		DA		AC		MT		Overall	
	Art	Mat	Art	Mat	Art	Mat	Art	Mat	Art	Mat
Neg Concl	94	97	63	37	76	25	76	85	77	61
Aff Concl	93	97	40	10	65	19	46	52	61	45

The mathematics group showed a significant effect on the MT inference, $p < 0.001$, but none on either MP, DA or AC.

To further explore these data, we adopted a similar procedure to that followed by Evans and Handley (1999). Participants' NCIs were subjected to an analysis of variance (ANOVA) with two within-participant factors (inference-type and negation-type), and one between-participants factor (group). The negation-type factor referred to whether the inference involved explicitly negated premises or implicitly negated premises. Given the finding that the negative conclusion effect was only significant on denial inferences, the inference-type factor was obtained by collapsing the four inferences into two categories: affirmative (MP and AC) and denial (DA and MT). Consequently the indices ranged from -4 to 4 , giving a reasonable approximation to the continuous distribution assumption of ANOVA. As predicted by the literature, there was a significant effect for inference-type, $F(1, 63) = 33.9$, $p < 0.001$, with a mean NCI for denial inferences of 1.75 compared to 0.28 for affirmative inferences. The mean NCIs for each group, for each inference-type are shown in Figure 3. There was no significant effect for negation-type, nor were there significant inference-type \times group, negation-type \times group or inference-type \times negation-type \times group interaction effects, all $F_s < 1$.

In sum, both the mathematics students and the arts students exhibited the negative conclusion effect: they were more likely to endorse denial inferences with negative conclusions than they were with affirmative conclusions. Furthermore, there were no significant differences between the degrees by which each group was affected.

3.3. THE AFFIRMATIVE PREMISE EFFECT.

To test for the affirmative premise effect we again followed the procedure used by Evans and Handley (1999), by calculating an *Affirmative Premise Index* (API). This was given by the number of inferences endorsed on arguments with affirmative premises minus the number of inferences endorsed on arguments with negative premises. Thus a participant who showed no affirmative premise effect would be expected to have an API of zero, and a

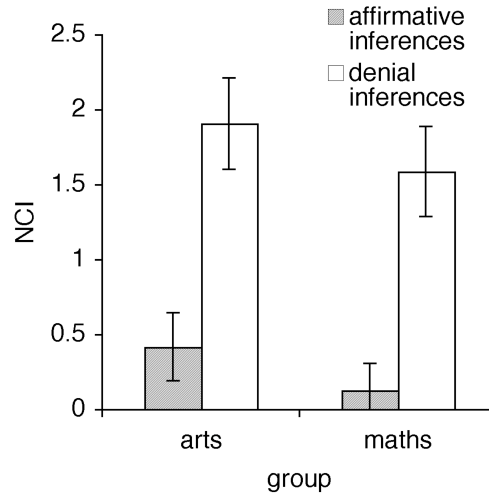


Figure 3. The mean NCIs for each group, for each inference-type. Error bars represent ± 1 SE of the mean.

participant who did show such an effect would be expected to have a positive API. A participant who exhibited the standard effect from the literature (of exhibiting a larger affirmative premise effect on inferences with implicitly negated premises) would be expected to have a higher API on inferences with implicitly negated premises than on inferences with explicitly negated premises. Equally, a participant who exhibited no such difference would be expected to have equal APIs for the two negation-types.

Table V shows the percentage of inferences made from affirmative premises and negative premises, for each of the two negation-types. For each group the significance of the affirmative premise effect was assessed by comparing the APIs to zero using a one-sample Wilcoxon signed ranks test.⁶ Overall, the arts group followed the pattern predicted by the literature and showed a large significant affirmative premise effect on implicitly negated premises, $p < 0.001$, and a significant, but smaller, effect on explicitly negated premises, $p = 0.002$. However, no significant effects were found for the mathematics group on either negation-type. Considering each inference separately showed significant effects for the arts group on MP (implicit), $p = 0.002$, AC (implicit), $p < 0.001$, DA (implicit), $p = 0.008$, and DA (explicit), $p = 0.002$. The mathematics group showed no significant effects on any inference.

The main analysis followed a slightly different route to that used for the NCIs. Due to Schroyens et al.'s (2001) observation that the affirmative

⁶ As with the NCI analysis, we set a more stringent α -value of 0.01 for this set of Wilcoxon signed-rank comparisons.

Table V. The percentage of inferences with affirmative and negative premises made by each group (Art – arts, Mat – mathematics), for each premise negation-type.

	MP		DA		AC		MT		Overall	
	Art	Mat	Art	Mat	Art	Mat	Art	Mat	Art	Mat
Explicitly negated premises										
Aff Prem	100	100	56	18	85	32	66	68	77	54
Neg Prem	94	100	41	13	66	15	59	68	65	49
Implicitly negated premises										
Aff Prem	100	100	60	15	88	31	74	71	81	49
Neg Prem	84	92	37	6	44	11	56	76	52	42

premise effect tends to be slightly stronger on invalid inferences, we replaced the inference-type factor with a validity factor. Consequently the APIs were subjected to an ANOVA with two within-participant factors (validity and negation-type), and one between-participants factor (group). As before, the negation-type factor referred to whether the inference involved premises which were explicitly or implicitly negated, whereas the new validity factor came from collapsing inferences into those which are valid (MP and MT) and those which are invalid (DA and AC). As with the NCI analysis, then, the indices ranged from -4 to 4 .

As predicted by the literature there was a significant effect for validity, $F(1, 63) = 22.5$, $p < 0.001$, with a mean API of 0.36 for valid inferences compared to 1.54 for invalid inferences. There was, however, no validity \times group interaction, $F < 1$. Importantly, there was a significant negation-type \times group interaction, $F(1, 63) = 4.04$, $p = 0.049$. This interaction is shown visually in Figure 4. The arts group showed a large difference in affirmative premise effects between the two negation-types, with a difference between the mean APIs of implicitly and explicitly negated premises of $+0.91$, $t(33) = 2.67$, $p = 0.012$. In contrast the difference for the mathematics group was only $+0.04$, $t(30) = 0.12$, NS. No significant validity \times negation-type, $F(1, 63) = 3.18$, NS, or validity \times negation-type \times group, $F < 1$, interaction effects were found.

In sum, the arts students in the sample exhibited the standard affirmative premise effect. They endorsed more inferences with affirmative premises than they did inferences with negated premises, with a larger effect when these negated premises were expressed implicitly compared to when they were expressed explicitly. In contrast, however, the mathematics group exhibited no effect on either the implicit or explicit negation-types.

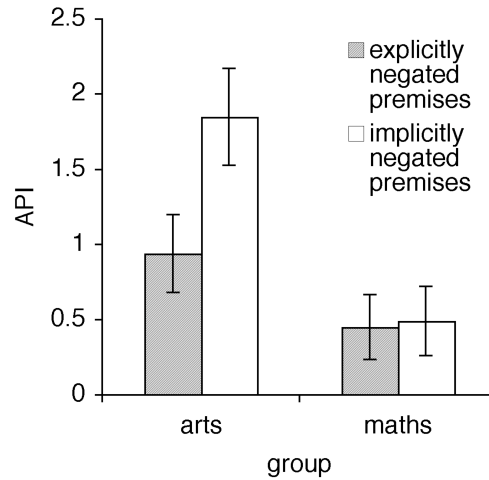


Figure 4. The mean APIs for each group, for each negation-type. Error bars represent ± 1 SE of the mean.

4. Discussion.

4.1. SUMMARY OF MAIN FINDINGS.

The main goals of this study were to (i) determine whether advanced mathematics students draw more normative inferences from abstract conditional statements than students who have not studied advanced mathematics; and, (ii) to investigate whether successful mathematics students exhibit the two main effects which have been found to interfere with normative logical behaviour when drawing conditional inferences; namely (a) the negative conclusion effect, and (b) the affirmative premise effect.

The picture which emerged was unexpectedly mixed. Our data showed that those students with experience of advanced mathematics did indeed make more normatively correct choices on an abstract conditional inference task than those without such experience. Interestingly, however, the only significant difference between the two groups came on the invalid AC and DA inferences. In short, the mathematics students were better than the arts students at rejecting invalid inferences, but not significantly better at endorsing valid inferences. In particular, this finding replicates and extends O'Brien's (1973) finding that many mathematics students have difficulty with MT inferences (in our study 69% and 61% of MT inferences were endorsed by the mathematics and arts groups respectively; in O'Brien's study, albeit with contextualised materials, the equivalent figure was 57%). We also replicated Durand-Guerrier's (2003) finding that O'Brien's 'child logic' is consistently used by very few students.

With respect to the negative conclusion effect, our data showed that both the groups tended to endorse more denial inferences with negative conclusions than they did denial inferences with positive conclusions. The difference between mean NCIs for denial and affirmative inferences was approximately equal for the two groups, suggesting that both were affected by the negative conclusion effect by similar degrees.

In contrast, there was a between-groups difference with regards to the affirmative premise effect. The arts group showed the standard effect: they endorsed more inferences with affirmative premises than they did inferences with negated premises, and the effect was larger when those negations were represented implicitly. The mathematics group, in contrast, showed no effect. Crucially, there was a significant negation-type \times group interaction for APIs, suggested that the affirmative premise effect did manifest itself differently for the two groups.⁷ In the remaining two sections of this discussion we consider issues surrounding the negative conclusion and affirmative premise effects, and suggest links between our results and recent theories on mathematical representations.

4.2. THE NEGATIVE CONCLUSION EFFECT.

The negative conclusion effect refers to the tendency of participants to endorse more inferences which lead to negative conclusions than those which lead to affirmative conclusions. As predicted by the literature, the arts group in our study exhibited this effect on denial inferences. But, despite making more normatively correct responses overall, the mathematics group also exhibited the effect on denial inferences, by a similar degree of magnitude to the arts students.

As discussed, one mechanism which has been proposed to account for the negative conclusion effect is related to the difficulty of making the deductive step from not-(not- p) to p . Consider, for example, the two inferences in Figure 5. The mental logic account of reasoning proposes that MT is an application of MP with a contradiction argument (Braine and O'Brien, 1998). Figure 5, supposing this process, shows the inferential steps required for each inference.⁸ The righthand inference (with a positive conclusion) requires the extra step of double negation compared to the lefthand inference (with a negative conclusion), perhaps making the inference harder to complete. It is of course well known that students have difficulty negating complex quantified

⁷ Note that this interaction cannot be accounted for by the mathematics group's tendency to endorse fewer invalid inferences than the arts group. If this were the primary reason for the negation-type \times group interaction, one would also expect a significant inference-type \times group interaction effect, whereas actually this effect did not approach significance, $F < 1$.

⁸ An essentially identical analysis (in the sense that it invokes a double negation process to account for the negative conclusion effect) can be made using the mental models theory of reasoning (Johnson-Laird, 2006).

<i>Rule:</i> If the letter is T then the number is 6.	<i>Rule:</i> If the letter is not T then the number is 6.
<i>Premise:</i> The number is not 6.	<i>Premise:</i> The number is not 6.
<i>Conclusion:</i> The letter is not T.	<i>Conclusion:</i> The letter is T.
<i>Inferential steps:</i>	<i>Inferential steps:</i>
Given: (i) ‘if T then 6’ and (ii) ‘not-6’.	Given: (i) ‘if not-T then 6’ and (ii) ‘not-6’.
Suppose (iii) T.	Suppose (iii) not-T.
Infer (iv) 6, by MP and (iii).	Infer (iv) 6, by MP and (iii).
Infer (v) 6 and not-6, by (ii) and (iv).	Infer (v) 6 and not-6, by (ii) and (iv).
Conclude not-T, by contradiction of (v), and (iii).	Conclude (vi) not-(not-T), by contradiction of (v), and (iii).
	Conclude T, by double negation of (vi).

Figure 5. Two MT inference tasks, one with an affirmative conclusion (right) and one with a negative conclusion (left), together with the steps needed to make the inference (according to mental logic theorists).

statements in mathematical contexts (e.g. Barnard, 1995; Dubinsky et al., 1988), but we are aware of no previous study which has found successful mathematics students having difficulty with the negation of abstract ‘not- p ’ statements, albeit when under the large cognitive load of holding in mind several previous inferential steps.

4.3. THE AFFIRMATIVE PREMISE EFFECT.

The affirmative premise effect refers to the phenomenon of participants endorsing more inferences with affirmative premises than negative premises. The effect tends to be stronger when negative premises are represented implicitly (i.e. as ‘8’ rather than ‘not-1’). In this paper we have reported evidence that arts students exhibited, as predicted by the literature, an increased effect for inferences with implicitly negated premises over explicitly negated premises; but, in contrast, we found that participants with a background in proof-based mathematics exhibited no effect at all. In this section we consider one account for this effect, and on its relation with advanced mathematical study.⁹

Evans and Handley (1999) suggested that the reason why the affirmative premise effect is substantially larger on inferences with implicitly negated premises is that participants see fewer implicitly negated premises as relevant to the conditional than they do explicitly negated premises, and so they

⁹ It should be emphasised that there is considerable debate about the mechanism behind the affirmative premise effect, and our goal here is not to intervene in disputes between reasoning theorists (interested readers can consult the discussions between Evans, 2002; Oaksford, 2002; and Yama, 2001).

<i>Rule:</i> If the letter is H then the number is 1.	<i>Rule:</i> If the letter is H then the number is 1.
<i>Premise:</i> The number is not 1.	<i>Premise:</i> The number is 8.
<i>Conclusion:</i> The letter is not H.	<i>Conclusion:</i> The letter is not H.

Figure 6. Two MT inference tasks, one with an explicitly negated premise (left) and one with an implicitly negated premise (right).

make fewer inferences. In short, participants fall at the first hurdle of the ‘two hurdles’ account. The crucial factor here appears to be the way premises in conditional inference tasks are represented. The nature of mathematical representations, and their influence on the learning of mathematics has received considerable attention in recent years, and in the discussion below we draw parallels between our work and contemporary studies on mathematical representations.

Lesh et al. (1987) introduced the notion of *transparent* and *opaque* representations of mathematical concepts. In their terms, an opaque representation emphasises some aspects of the concept, whilst de-emphasising others; whereas, in contrast, a transparent representation neither adds nor takes away meaning from the underlying concept. Zazkis and Gadowsky (2001) argued that a single representation may be transparent with respect to certain properties, but opaque with respect to others. For example, one way of representing the number 1007 is $1000 + 7$, another way is 53×19 . The first representation emphasises that 1007 is larger than 1000, but de-emphasises that it is divisible by 53 (it is transparent with respect to magnitude, but opaque with respect to divisibility). The second representation has the reverse property (it is transparent with respect to divisibility, but opaque with respect to magnitude). Zazkis and Liljedahl (2004) argued that the reason students have considerable difficulties with understanding prime numbers and prime factorisation is because of the lack of a transparent representation system for primes, and similar observations can be made of higher level mathematics. In the domain of knot theory, for example, the Tait Conjectures remained unproved for much of the twentieth century, but upon the discovery of new and more transparent ways of representing knots – the Jones and HOMFLY polynomials – the conjectures were quickly solved.

Kirshner and Awtry (2004) noted that representation systems for some mathematical concepts may be opaque in a slightly different way: rather than de-emphasising certain valid properties, they may emphasise other *invalid* properties. Kirshner and Awtry suggested that many of the widely documented ‘mal-rules’ applied by students when making algebraic manipulations are a consequence of such effects. For example, both the mal-rule $a^{m+n} = a^m + a^n$, and the actual rule $a(m+n) = am + an$ are salient: the left and right sides of the equations appear naturally similar. Kirshner and Awtry

found that students were more likely to accurately complete transformations if they had this saliency property. No such difference, however, was found for students who has been taught algebra with a representation system which eliminated these saliency features. Complementary results were published by Landy and Goldstone (2007), who found that the internal spacing of algebraic expressions can affect their opaqueness/transparency.

The affirmative premise effect could, then, be reconceptualised using the terminology adopted by Lesh et al. (1987). In the example given in Figure 6, the notation ‘8’ is an opaque representation for the concept not- q ; in contrast, the notation ‘not 1’ is a transparent representation, as it shares salient features with the conditional ‘if p then q ’: the presence of 1. In short, Evans and Handley’s (1999) account of the affirmative premise effect suggests that participants struggle to see through the opaqueness of the notation used in the implicitly negated component of the task, in exactly the same way as the students studied by Kirshner and Awtry (2004), Lesh et al., and Zazkis and Liljedahl (2004) struggled to see through the opaqueness of the notations used in their tasks. However, the crucial difference between our study, and those of other researchers who have looked at opaque and transparent representations, is that unlike the arts students, the mathematics students in our sample (all undergraduates who had been highly successful at school level mathematics) were apparently *not* affected by the opaqueness of the representation of implicitly negated premises: they exhibited no affirmative negation effect on either negation-type.

All these disparate studies point in the same direction: towards the speculative hypothesis that one characteristic of successful mathematical thinking is an ability to override the attentional biases that are generated by opaque representations. This study suggests that mathematics students are not affected by the opaqueness of implicit negations on an abstract conditional inference task. Similarly Kirshner and Awtry (2004), Lesh et al. (1987) and Zazkis and Liljedahl (2004) have shown that one of the primary difficulties weak students have with mathematics is related to exactly these types of opaque representations. It seems reasonable, therefore, to speculate that the ability to ‘see through’ opaque representations is a characteristic of mathematically successful students. Expert/novice differences in the deployment of attention whilst engaged in mathematical activities would appear to be a fruitful area for future research.

5. Conclusion.

Many curriculum bodies and commentators have supported a form of the theory of formal discipline by expressing the view that studying mathematics facilitates abstract logical reasoning. Despite the sceptical views of early

psychologists towards this position (e.g. Thorndike and Woodworth, 1901), Lehman and Nisbett (1990) found a slight correlation between the number of mathematics courses taken by science and humanities students during their degree and their improvement on a test of conditional reasoning. To investigate this view, we directly compared the inferences drawn by mathematics and arts undergraduates from abstract conditional statements. The overall picture we found was mixed. Successful mathematics students endorsed fewer invalid inferences than a control group of arts students, but they did not endorse more valid inferences. We paid particular attention to two effects which have been found to interfere with normative logical behaviour: the negative conclusion effect and the affirmative premise effect. It was found that both the arts and mathematics students in our sample endorsed significantly more inferences which led to negative conclusions than they did inferences which led to affirmative conclusions. In contrast, only the arts group exhibited the affirmative premise effect: they endorsed more inferences with affirmative premises than they did inferences with negative premises. This was not the case with the mathematics group.

One unresolved question is related to the *cause* of these between-groups differences. Whilst we have found a correlation between the study of advanced mathematics and a different pattern of responses to an abstract conditional inference task, we have not established the causal relationship posited by the formal discipline view. There are two reasonable hypotheses. Perhaps, as suggested by those who support a form of the theory of formal discipline (Oakley, 1946; Smith, 2004), exposure to the study of higher level mathematics brings about the development of certain logical skills: in particular the ability to recognise invalid inferences, and to see through opaque representations. An alternative hypothesis, however, is that there is a subset of the population who are naturally skilled in these areas, and that it is this subset who are disproportionately filtered into studying mathematics at advanced levels. Although our cross-sectional data do not allow us to distinguish between these hypotheses, in view of Lehman and Nisbett's (1990) findings we feel there is more support for the formal discipline view, albeit with important caveats. In particular it is clear that there is no simplistic relationship between advanced mathematical study and abstract logical reasoning skills: the idea that mathematicians are simply 'better' at abstract logic is too simplistic (indeed, there are some areas where they are not, apparently, any better). But, equally, we have also demonstrated that theories which posit no relationship are also too simplistic. We found that in some crucial aspects of conditional inferential reasoning mathematics undergraduates behave more normatively than arts undergraduates, whereas in some other aspects they do not. Determining whether these differences are developed, and, if they are, at which curriculum stage they develop, would be a valuable goal for future research.

Acknowledgements

We would like to thank Gary Davis, Paola Iannone and Keith Weber for their helpful comments on earlier drafts of this work.

References

- Barnard, A. D.: 1995, 'The impact of 'meaning' on students' ability to negate statements'. In: L. Meira and D. Carraher (eds.): *Proceedings of the 19th International Conference on the Psychology of Mathematics Education*, Vol. 2. Recife, Brazil, pp. 3–10, IGPME.
- Braine, M. D. S. and D. P. O'Brien: 1998, *Mental logic*. Mahwah, NJ: Erlbaum.
- Cheng, P. W., K. J. Holyoak, R. E. Nisbett, and L. M. Oliver: 1986, 'Pragmatic versus Syntactic Approaches to Training Deductive Reasoning'. *Cognitive Psychology* **18**, 293–328.
- Damarin, S. K.: 1977, 'The Interpretation of Statements in Standard Logical Form by Preservice Elementary Teachers'. *Journal for Research in Mathematics Education* **8**, 123–131.
- Davis, C.: 1850/1970, 'The Logic and Utility of Mathematics'. In: J. K. Bidwell and R. G. Clason (eds.): *Readings in the History of Mathematics Education*. Washington DC: NCTM, pp. 39–62.
- Dubinsky, E., F. Elterman, and C. Gong: 1988, 'The student's construction of quantification'. *For the Learning of Mathematics* **8**(2), 44–51.
- Durand-Guerrier, V.: 2003, 'Which notion of implication is the right one? From logical considerations to a didactic perspective'. *Educational Studies in Mathematics* **53**, 5–34.
- Eisenberg, T. A. and R. L. McGinty: 1974, 'On Comparing Error Patterns and the Effect of Maturation in a Unit on Sentential Logic'. *Journal for Research in Mathematics Education* **5**, 225–237.
- Evans, J. St. B. T.: 2002, 'Matching bias and set sizes: A discussion of Yama (2001)'. *Thinking and Reasoning* **8**, 153–163.
- Evans, J. St. B. T.: 2007, *Hypothetical Thinking: Dual Processes in Reasoning and Judgement*. Hove, UK: Psychology Press.
- Evans, J. St. B. T., J. Clibbens, and B. Rood: 1995, 'Bias in conditional inference: Implications for mental models and mental logic'. *Quarterly Journal of Experimental Psychology* **48A**, 644–670.
- Evans, J. St. B. T. and S. J. Handley: 1999, 'The Role of Negation in Conditional Inference'. *Quarterly Journal of Experimental Psychology* **52A**, 739–769.
- Hoyles, C. and D. Küchemann: 2002, 'Students' Understanding of Logical Implication'. *Educational Studies in Mathematics* **51**, 193–223.
- Johnson, D. L.: 1998, *Elements of Logic via Numbers and Sets*. London: Springer-Verlag.
- Johnson-Laird, P. N.: 2006, *How we reason*. Oxford: OUP.
- Kirshner, D. and T. Awtry: 2004, 'Visual salience of algebraic transformations'. *Journal for Research in Mathematics Education* **35**, 224–257.
- Landy, D. and R. Goldstone: 2007, 'How abstract is symbolic thought?'. *Journal of Experimental Psychology: Learning, Memory and Cognition* **33**, 720–733.
- Lehman, D. R. and R. E. Nisbett: 1990, 'A Longitudinal Study of the Effects of Undergraduate Training on Reasoning'. *Developmental Psychology* **26**, 952–960.
- Lesh, R., M. Behr, and T. Post: 1987, 'Rational number relations and proportions'. In: C. Janiver (ed.): *Problems of representations in the teaching and learning of mathematics*. Hillsdale, N.J.: Lawrence Erlbaum, pp. 41–58.

- Newstead, S. E., M. Charles Ellis, J. St. B. T. Evans, and I. Dennis: 1997, 'Conditional Reasoning with Realistic Material'. *Thinking and Reasoning* **3**(1), 49–76.
- Nisbett, R. E. and T. D. Wilson: 1977, 'Telling more than we can know: Verbal reports on mental processes'. *Psychological Review* **84**, 231–295.
- Oakley, C. O.: 1946, 'Mathematics'. *American Mathematical Monthly* **56**, 19.
- Oaksford, M.: 2002, 'Contrast classes and matching bias as explanations of the effects of negation on conditional reasoning'. *Thinking and Reasoning* **8**, 135–151.
- Oaksford, M. and N. Chater: 2007, *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. Oxford: OUP.
- Oaksford, M., N. Chater, and J. Larkin: 2000, 'Probabilities and polarity biases in conditional inference'. *Journal of Experimental Psychology: Learning, Memory and Cognition* **26**, 883–899.
- O'Brien, T. C.: 1972, 'Logical Thinking in Adolescents'. *Educational Studies in Mathematics* **4**, 401–428.
- O'Brien, T. C.: 1973, 'Logical Thinking in College Students'. *Educational Studies in Mathematics* **5**, 71–79.
- O'Brien, T. C., B. J. Shapiro, and N. C. Reali: 1971, 'Logical Thinking – Language and Context'. *Educational Studies in Mathematics* **4**, 201–219.
- Pollard, P. and J. St. B. T. Evans: 1980, 'The influence of logic on conditional reasoning performance'. *Quarterly Journal of Experimental Psychology* **32**, 605–624.
- QAA: 2002, 'Mathematics, statistics and operational research subject benchmark standards'. Online article [accessed 15/07/2005]: <http://www.qaa.ac.uk/academicinfrastructure/benchmark/honours/mathematics.pdf>.
- Schooler, J. W., S. Ohlsson, and K. Brooks: 1993, 'Thoughts beyond words: When language overshadows insight'. *Journal of Experimental Psychology: General* **122**, 166–183.
- Schroyens, W., W. Schaeken, and G. d'Ydewalle: 2001, 'The processing on negations in conditional reasoning: A meta-analytical case study in mental model and/or mental logic theory'. *Thinking and Reasoning* **7**, 121–172.
- Schroyens, W., W. Schaeken, W. Fias, and G. d'Ydewalle: 2000, 'Heuristic and Analytic Processes in Propositional Reasoning With Negatives'. *Journal of Experimental Psychology: Learning, Memory and Cognition* **26**, 1713–1734.
- Smith, A.: 2004, *Making Mathematics Count: The report of Professor Adrian Smith's Inquiry into Post-14 Mathematics Education*. London: The Stationery Office.
- Stanic, G. M. A.: 1986, 'The Growing Crisis in Mathematics Education in the Early Twentieth Century'. *Journal for Research in Mathematics Education* **17**, 190–205.
- Stylianides, A. J., G. J. Stylianides, and G. N. Philippou: 2004, 'Undergraduate Students' understanding of the contraposition equivalence rule in symbolic and verbal contexts'. *Educational Studies in Mathematics* **55**, 133–162.
- Thorndike, E. L.: 1924, 'Mental discipline in high school studies'. *Journal of Educational Psychology* **15**, 1–22.
- Thorndike, E. L. and R. S. Woodworth: 1901, 'The influence of improvement in one mental function upon the efficiency of other functions'. *Psychological Review* **8**, 247–261.
- Wason, P. C. and J. St. B. T. Evans: 1975, 'Dual Processes in Reasoning?'. *Cognition* **3**, 141–154.
- Yama, H.: 2001, 'Matching versus optimal data selection in the Wason Selection Task'. *Thinking and Reasoning* **7**, 295–311.
- Zazkis, R. and K. Gadowsky: 2001, 'Attending to transparent features of opaque representations of natural numbers'. In: A. Cuoco (ed.): *The roles of representation in school mathematics*. Reston, VA: NCTM, pp. 146–165.
- Zazkis, R. and P. Liljedahl: 2004, 'Understanding primes: The role of representation'. *Journal for Research in Mathematics Education* **35**, 164–186.

Address for Offprints:

Matthew Inglis
Learning Sciences Research Institute
University of Nottingham
Nottingham, United Kingdom.
e-mail: matthew.inglis@nottingham.ac.uk